International Journal of Educational Research 114 (2022) 102000

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

The correspondence between PISA performance and school achievement in Finland☆

Jonna Pulkkinen *, Juhani Rautopuro

Finnish Institute for Educational Research, University of Jyväskylä, Finland

ARTICLE INFO

Keywords:
PISA
Achievement
Grades
Basic education
Gender
Socio-economic status

ABSTRACT

In the Programme for International Student Assessment (PISA), the key knowledge and skills needed in modern society are assessed. So far, however, little is known on the relationship between PISA performance and school achievement. Using Finnish PISA 2015 and 2018 data combined with the national register of education records, we studied whether there is correspondence between PISA performance and school achievement. The results showed that the correspondence between PISA performance and school achievement is moderate. The PISA proficiency scores correlated not only with the corresponding grades but also with the grades of other theoretical subjects, indicating that the PISA test assesses a wide range of school achievement. PISA proficiency was related to grades even after controlling for gender and SES.

1. Introduction

In the Programme for International Student Assessment (PISA), 15-year-old students' performance in reading, mathematics, and science is studied in every three years. PISA is not based on national curricula. Instead, the aim is to assess the key knowledge and skills that are essential for participation in modern society and to examine how students can apply knowledge in different settings (OECD, 2016). Schleicher (2007) claimed that although the PISA test does not assess all the competencies needed in modern society, performance on the PISA test predicts the future success of students. To ensure the international validity of the PISA test, all participant countries or economies collaborate with the organisation for Economic Co-operation and Development (OECD) on developing the PISA assessment, and the assessment framework of each domain is developed by expert groups (OECD, 2016).

However, PISA has also been criticized. For example, Dohn (2007) stated that PISA can assess only the knowledge and skills needed in the PISA test situation, not the knowledge and skills needed in real-life situations. Moreover, it has been argued that there are problems concerning the measurement of non-cognitive variables, the comparability of achievement and background variables across countries, and reporting and interpreting the results (e.g., Hopfenbeck & Maul, 2011; Kivinen & Hedman, 2017; Rutkowski & Rutkowski, 2010, 2016; Rutkowski & Rutkowski, 2013, 2018).

Nevertheless, many studies have utilized PISA data as secondary data, such as to study educational inequalities (Hopfenbeck et al., 2018). Although PISA data has been used as secondary data in many studies, few studies have combined the PISA data with other student data to focus on the relationship between PISA performance and students' school achievement. In this study, we examine how PISA performance predicts students' school achievement, as measured by school grades. We use the Finnish PISA data combined with

☆ This work was supported by the Strategic Research Council, Academy of Finland under Grant number #314314.

* Corresponding author.
E-mail address: jonna.pulkkinen@jyu.fi (J. Pulkkinen).

https://doi.org/10.1016/j.ijer.2022.102000
Received 15 July 2021; Received in revised form 12 May 2022; Accepted 14 May 2022
Available online 25 May 2022

the national register data of education records that include students' grades at the end of basic education. Thus, the data are unique. Our research questions are as follows:

1 How strong is the correlation between PISA proficiency scores in reading, mathematics and science, and school achievement?
2 To what extent do PISA proficiency levels in reading, mathematics, and science predict the corresponding school grades in each subject?

2. The PISA test as a measure of student achievement

The PISA test assesses student proficiency in the core school subjects (reading, mathematics, and science) but is not directly based on the curricula of participating nations (OECD, 2019a). The PISA test puts great emphasis on problem solving and the student's ability to apply acquired knowledge, with the aim of assessing the knowledge and skills needed in modern society (OECD, 2019a). Consequently, determining the PISA test's validity in predicting student achievement in education and the labor market is of great interest.

The relationship between performance on the PISA test and achievement in education has been studied by using both grades and transition to further education as an indicator of achievement. An example of a study that has focused on the relationship between PISA performance and grades is that of Fischbach et al. (2013). In their study, Luxembourg's PISA 2006 data and official school records of grades were used. According to the authors, PISA proficiency scores did correlate with corresponding grades, and the effect sizes of the correlations were medium. Fischbach et al. (2013) also found that PISA proficiency scores correlated with non-corresponding grades, which, according to them, showed that the discriminant predictive power of PISA performance is weak.

Matějů & Smith (2015), instead, utilized Czech PISA 2003 data to examine the potential gender gap in PISA performance and students' self-reported grades and educational aspirations. They found differences between girls and boys in the relationship between PISA performance and school achievement. With the same PISA performance, girls received better grades in Czech language and in mathematics. Moreover, girls' chance to receive top grades was greater than boys' chance even when PISA performance and socio-economic status (SES) was controlled for (Matějů & Smith, 2015). Harju-Luukkainen et al. (2016) found regional differences in the correspondence between PISA proficiency scores and school grades. They studied the relationship between PISA reading scores and students' self-reported grades using the Finnish PISA 2009 data. The study showed that students performed mainly better on the PISA test than could be expected on the basis of their grades although the opposite situation occurred in some areas. Harju-Luukkainen et al. (2016) concluded that this indicates that grading practices are adjusted to the general competence level of the class or school.

The relationship between PISA proficiency scores and educational achievement has also been studied by following PISA students' transition to further education. For example, Bertschy et al. (2009) examined Swiss PISA 2000 students' transition to vocational training and the labor market. According to their study, higher PISA scores in reading influenced the probability of receiving a more intellectually demanding vocational training. Murdoch et al. (2011), instead, studied Canadian PISA 2000 students' transition to higher education and graduation or dropping out within higher education. They found that PISA scores in reading influenced the access to higher education independently of previous schooling and social factors (e.g., the gender or occupation and level of education of each parent). In addition, their study showed that PISA scores had a greater impact on the access to education than on persistence within higher education.

3. Grades as a measure of students' achievement

Criteria for school grades are usually described in the national curricula. Although the composition of grades may vary in different nations, they are rarely based on a single test and do not simply reflect student knowledge. As presented above, the studies show that there is a relationship between grades and achievement on the PISA test but that the relationship is quite weak. The relationship between grades and achievement on standardized tests has also been studied using national tests. For example, Klapp Lekholm & Cliffordson (2008, 2009) studied Swedish ninth-grade students' achievement in national tests and their school grades in Swedish, English, and mathematics using register data. The focus of their research was to examine whether grades are affected by both achievement and other student characteristics. According to their studies, the variance in grades was mostly explained by achievement on national tests. However, they found that there is also a common grade dimension. The study by Klapp Lekholm & Cliffordson (2009) showed that, in addition to achievement, grades are affected by other factors such as self-perceptions of competence, motivation, and support from parents. Similarly, Bowers (2011) argued based on his data on American high school students that grades not only describe academic knowledge but also assess non-cognitive behavior. The common school grade dimension has also been shown to predict students' educational success in upper secondary education (Thorsen & Cliffordson, 2012).

In previous studies, some student characteristics, such as gender and SES, have been found to be related to grades. In their meta-analysis on gender differences in educational achievement, Voyer & Voyer (2014) showed that girls receive better grades than boys. According to the authors, the above-mentioned difference was largest in language and smallest in mathematics. Likewise, Klapp Lekholm & Cliffordson (2008, 2009) found gender differences in grades in girls' favor. Moreover, according to Klapp Lekholm & Cliffordson (2008) grades are influenced by family educational background so that the grades of students with a lower educational background were slightly higher.

Results similar to those presented above were achieved in Finland, where the national assessment of learning outcomes at the end of basic education has been reported by school subject since the beginning of the 21st century. The discrepancies between the grades given to students and the performance they show in assessments of learning outcomes, such as in mathematics and in their mother tongue and literature, were detected at the very beginning (Lappalainen, 2004, 2006; Mattila, 2002). Although the correlation

between grade and performance seemed to be rather high in these assessments, the performance of students who received the same grades varied. This phenomenon has remained the same over the years (Hildén & Rautopuro, 2017). In mathematics, for example, boys and students whose parents had attained a higher educational level received lower grades with relation to their performance.

Moreover, Finnish studies have shown that there are discrepancies in school grades between schools and even within the schools. At worst, students with a similar average performance on the assessment could have two units difference in grades depending on the school (e.g., Hilden & Rautopuro, 2017; Hotulainen et al., 2016; Julin & Rautopuro, 2016). One interesting finding is that schools that perform better than average tend to have a more rigid grading culture (e.g., Mattila, 2002). Thus, the results of earlier studies show that grades are not only related to student knowledge or assessment performance but are affected by other student characteristics.

4. Education and evaluation system in Finland

This study was conducted in Finland, where the education system is based on an equal and equitable basic education. The comprehensive schools do not select their students, and the aim of basic education is to offer all students an equal opportunity to receive education irrespective of students' background (Basic Education Act, 1998). The goal of the National Core Curriculum (NCC) is to ensure equal and fair access to upper secondary education. In Finland, students apply to upper secondary education at the end of basic education. Upper secondary education usually lasts three years and includes two different tracks: general or vocational upper secondary education (Eurydice, 2021). Access to a desirable general upper secondary school or field of vocational education is mainly based on students' grades in the final basic education assessment. Consequently, the national comparability of grades at the end of basic education is of critical importance in Finland since grades determine students' access to upper secondary education (Ouakrim-Soivio et al., 2018). In Sweden, for example, where upper secondary education also lasts three years and is divided into different vocational or higher education preparatory programs, the student must pass a specific number of subjects to be eligible for upper secondary education (Eurydice, 2021).

In Finland, there are no high-stakes standardized national tests during or at the end of basic education. Therefore, the teachers are responsible for giving grades to students, and the NCC defines the baseline for the numerical grade in each subject. This should serve as a national guideline for teachers. According to the NCC, the student assessment focuses on the students' learning, working skills, and behavior. Assessment of working skills is a part of assessment and grade formulation in different subjects. Behavior is assessed as a separate entity and does not affect a subject's grade (Finnish National Agency for Education, 2004).

In Finnish basic education, students are assessed on a scale from 4 (failed) to 10 (excellent). Higher grades should reflect better performance in different school subjects. The assessment focuses on how a given student has achieved the objectives of the curriculum (Basic Education Decree 852, 1998). Until 2020, the assessment criteria were given only for school grade 8 (good competence). At the end of 2020, the new criteria for the final assessment of basic education were introduced, and these criteria came into effect in August of 2021 (Finnish National Agency for Education, 2020). These new criteria will set the criteria for school grades 5 (adequate), 7 (satisfactory), 8 (good), and 9 (very good).

Instead of high-stakes standardized tests, Finland offers some low-stakes national domestic assessments in different school subjects. They are sample-based, and they are conducted according to the evaluation plan given by the Finnish Ministry of Education and Culture. The aim of the national assessments is mainly to measure how well the aims of the NCC have been met and to examine the equity of learning outcomes (e.g., Jakku-Sihvonen, 2013). From an equitable point of view, the differences in learning outcomes between sub-groups (e.g., grouped by gender, school's language of instruction, or region) are studied more closely. Even if differences between individuals in learning outcomes are acceptable, systematic differences between sub-groups may present a possible threat to the equity of education (Jakku-Sihvonen & Kuusela, 2002, p. 7). In contrast to Finland, standardized national tests are used in other Nordic Region countries, such as Denmark, Estonia, Norway, and Sweden (Volmari, 2019). The primary aim of these tests is to ensure that all students are assessed equivalently. For example, in Sweden, there are tests for three core subjects (Swedish, mathematics, and English [foreign language]) in grades 6 and 9 consisting of both oral and written components (Skolverket, 2022).

Similar to other Nordic countries, Finland also participates in many international comparative assessment studies, such as the PISA, the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS). In the PISA study, Finnish students have succeeded very well. Although the trend in PISA results has been declining in Finland since 2009, the performance of Finnish students is still clearly above the average performance of other OECD countries (OECD, 2019b). The TIMSS and PIRLS, conducted by the International Association for the Evaluation of Educational Achievement (IEA), are, unlike the PISA study, based on national curricula, and they also show that the Finnish students perform well in international comparison (Mullis et al., 2017, 2020).

5. Method

5.1. Sample

In this study, we used the Finnish PISA 2015 and 2018 data combined with the national register of education records. These unique data enabled us to link the students' performance on the PISA test and their grades in their basic education certificate. The students participating in the PISA study are 15-year-olds. In Finland, most of the PISA students are in the ninth grade, which means that they are at the last grade level of basic education. In Finland, 5882 students participated in the PISA 2015 study. In 2018, the Finnish basic sample of students in the PISA test was 5649 students. Moreover, in Finland, there was a subsample of students ($n = 1954$) who participated in the assessment of financial literacy that was offered as an option in the PISA 2018 study (OECD, 2019b). This subsample

has been excluded from this study. In Finland, the test languages of the PISA test are Finnish and Swedish. The largest share of the students is tested in Finnish. Of all the PISA students, 6% ($n = 348$) were tested in Swedish in 2015 and 7% ($n = 386$) in 2018.

The data of this study include those Finnish PISA students who were in grade 9 and applied to upper secondary education through the joint application system in that spring they participated in the PISA study. The information on students' applications to upper secondary education is collected into the national register of education records. In addition to information on students' educational choices and results of application, the register includes information on students' grades from the basic education certificate (Act on the National Registers of Education Records, Qualifications & Degrees, 2017). The Finnish National Agency for Education has the responsibility for maintaining the register. The data from the year 2015 include 5015 students (representing 85.3% of all Finnish PISA 2015 students and 98.3% of the ninth-graders in the Finnish PISA 2015 sample). Of these students, 50.4% were boys as opposed to 51.3% of all Finnish PISA 2015 students. The data from the year 2018, instead, include 4794 students (representing 84.9% of all Finnish PISA 2018 students and 98.0% of the ninth-graders in the Finnish PISA 2018 sample), of whom 49.7% were boys (of all Finnish PISA 2018 students 50.9% were boys).

5.2. Measures

For the analysis, we used the PISA proficiency scores and levels in reading, mathematics, and science. In each cycle, one of these domains is a major domain. Science was as a major domain in the PISA 2015 study and reading was in the PISA 2018 study. In the PISA test, reading refers to "students' ability to understand, use, evaluate, reflect on and engage with text to achieve their purposes" (OECD, 2019a, p. 15). The assessment of reading focuses on students' proficiency in locating information, understanding text, and evaluating and reflecting on text (OECD, 2019a). The PISA test in mathematics assesses student proficiency in formulating, employing, and interpreting mathematics, whereas in science, the test assesses student proficiency in explaining phenomena and interpreting data scientifically as well as in evaluating and designing scientific inquiry (OECD, 2019a).

PISA proficiency scores for each domain were initially (in PISA 2000 study) reported so that the OECD average score was 500, and the standard deviation was 100 (OECD, 2019a). After that, the scores were scaled so that they would be comparable in different cycles, meaning that average scores are around 500 (OECD, 2019a). In 2015, the mean score of OECD countries was 493 in reading, 490 in mathematics, and 493 in science, whereas the mean scores of the Finnish students were 526, 511, and 531, respectively (OECD, 2016).

Table 1
Descriptive statistics for the variables used in this study.

	PISA 2015 ($N = 5015$) % (SE)		PISA 2018 ($N = 4794$) % (SE)	
Gender				
Girl	49.6 (0.55)		50.3 (0.54)	
Boy	50.4 (0.55)		49.7 (0.54)	
PISA proficiency level in reading				
Level 1	8.5 (0.63)		11.0 (0.59)	
Level 2	16.8 (0.87)		18.2 (0.75)	
Level 3	30.0 (0.99)		28.4 (0.84)	
Level 4	29.7 (1.04)		26.9 (0.82)	
Level 5	12.8 (0.69)		12.9 (0.79)	
Level 6	2.2 (0.28)		2.6 (0.41)	
PISA proficiency level in mathematics				
Level 1	10.8 (0.74)		12.0 (0.64)	
Level 2	21.3 (0.93)		21.8 (0.98)	
Level 3	30.2 (0.92)		30.1 (1.00)	
Level 4	25.3 (0.93)		24.1 (0.87)	
Level 5	10.0 (0.73)		10.0 (0.62)	
Level 6	2.4 (0.32)		2.0 (0.32)	
PISA proficiency level in science				
Level 1	9.1 (0.61)		10.5 (0.65)	
Level 2	18.3 (0.77)		19.9 (0.78)	
Level 3	29.7 (0.81)		29.7 (0.85)	
Level 4	27.5 (0.85)		26.5 (0.88)	
Level 5	12.7 (0.63)		11.4 (0.72)	
Level 6	2.6 (0.31)		2.0 (0.38)	
	M (SE)	SD (SE)	M (SE)	SD (SE)
PISA reading scores	535 (2.40)	89 (1.43)	529 (2.15)	96 (1.20)
PISA mathematics scores	518 (2.17)	79 (1.18)	514 (1.93)	79 (1.19)
PISA science scores	538 (2.25)	93 (1.24)	530 (2.54)	93 (1.23)
Grade in mother tongue and literature	7.92 (0.03)	1.18 (0.01)	7.96 (0.03)	1.18 (0.01)
Grade in mathematics	7.75 (0.03)	1.38 (0.02)	7.76 (0.03)	1.42 (0.01)
Mean of science-related grades	7.84 (0.03)	1.18 (0.01)	7.90 (0.03)	1.20 (0.01)
GPA of theoretical subjects	7.89 (0.03)	1.08 (0.01)	7.96 (0.02)	1.10 (0.01)
SES (ESCS index)	0.28 (0.02)	0.74 (0.01)	0.33 (0.02)	0.78 (0.01)

Note: Numbers are weighted values.

In 2018, the mean scores of OECD countries were 487 (reading), 489 (mathematics), and 489 (science), and the corresponding scores for the Finnish students were 520, 507, and 522, respectively (OECD, 2019b). The means and standard deviations of the PISA proficiency scores for the samples used in this study are presented in Table 1. In our samples, the mean scores were slightly higher than the average scores in Finland.

In addition to mean scores, PISA proficiency is described as levels. The levels define what kind of knowledge and skills a student needs to complete tasks at each level, and the range of each level is about 80 points (OECD, 2019b). In 2015, there were seven levels for reading and science and six levels for mathematics (OECD, 2016). In 2018, the proficiency scores were divided into eight levels in reading, seven levels in science, and six levels in mathematics (OECD, 2019b). In each domain, the highest level is Level 6. In reading and science, Level 1 is divided into two or three levels. In this study, we have combined different levels of Level 1 so that in the analysis we have six levels for each domain. Table 1 shows the proportion of students at these six levels.

Students' grades were obtained from the national register of education records. From the register, we used students' grades in the subjects that correspond to domains assessed in the PISA test (i.e., mother tongue and literature, mathematics, physics, chemistry, biology, and geography) and in other theoretical subjects (i.e., history, civics, religion or ethics, health education, foreign language, and second national language). The Finnish students participating in the PISA study in 2015 and 2018 studied according to the National Core Curriculum for Basic Education 2004. In mother tongue and literature, learning objectives (on which the final assessment is based) include students' interaction skills, use and interpretation of different kinds of text, production of text, and relation with language and literature (Finnish National Agency for Education, 2004). According to the NCC, mathematics learning objectives include thinking skills and methods, number and mathematical operations, algebra, functions, geometry, and probability and statistics. In addition to subject content knowledge, an essential learning objective of all science-related subjects is to develop students' scientific thinking and inquiry skills (Finnish National Agency for Education, 2004).

In the register, students can have grades in these subjects both for compulsory and optional syllabi. In this study, the grades for compulsory syllabi were used. The scale of the grades is as follows: 4 (failed), 5 (adequate), 6 (moderate), 7 (satisfactory), 8 (good), 9 (very good), and 10 (excellent). In the analysis, the mean variable of four science-related school grades (i.e., physics, chemistry, biology, and geography) was used. The Cronbach's alpha for this mean variable was 0.94 both in the PISA 2015 data and in the PISA 2018 data. Additionally, we calculated the grade point average (GPA) for all theoretical subjects. For the students of this study, the average school grades in mother tongue and literature, mathematics, and science-related subjects were about 8 (see Table 1).

The PISA index of economic, social, and cultural status (ESCS) was used as a measure of students' SES. The index includes components that describe students' family background. These are parents' highest level of education, parents' highest occupational status, and an index of home possessions. The latter describes the availability of household possessions and books at home. The ESCS index is scaled so that a value of 0 describes an average OECD student and that the standard deviation across OECD countries is 1 (OECD, 2019c, p. 52). In Finland, the mean of the ESCS index was 0.25 in 2015 (OECD, 2016) and 0.30 in 2018 (OECD, 2019c). The means of the ESCS index in the data used in this study are presented in Table 1.

5.3. Data analysis

First, to analyze the relationship between PISA proficiency scores and school achievement, we computed Pearson product moment correlations. Second, we conducted an ordinary least-squares linear regression analysis separately for each domain. In the regression analysis, we used PISA proficiency level instead of proficiency score as an independent variable, as proficiency levels are more understandable than scores. We executed two different models. In Model 1, the grade was predicted by only the PISA proficiency level. Since earlier studies show that student background variables, such as gender and SES, might affect school achievement, in Model 2, gender and SES were added to the model to control for these variables. All the analyses were run on both the PISA 2015 and PISA 2018 data.

In PISA, a two-stage sampling design is used, so student final weights and 80 replicate weights for computing the standard errors need to be used in the analyses (OECD, 2009). Moreover, when students' achievement is included in the analyses, plausible values (PVs) are used. PVs are random values that are based on the probability distribution of a student's ability and illustrate a range of possible values that a student's ability can obtain (OECD, 2009). The PISA 2015 and PISA 2018 data contain 10 PVs; consequently, analyses have to be performed on each PV, and these results are then aggregated. In the analyses, the above methodologies used in

Table 2
Correlations between PISA proficiency scores and school achievement.

PISA 2015	**PISA reading scores**	**PISA mathematics scores**	**PISA science scores**
Grade in mother tongue and literature	.58*** (0.01)	.49*** (0.02)	.56*** (0.01)
Grade in mathematics	.57*** (0.01)	.58*** (0.01)	.62*** (0.01)
Mean of science-related grades (physics, chemistry, biology, and geography)	.62*** (0.01)	.60*** (0.01)	.66*** (0.01)
GPA of theoretical subjects	.65*** (0.01)	.61*** (0.01)	.67*** (0.01)
PISA 2018	**PISA reading scores**	**PISA mathematics scores**	**PISA science scores**
Grade in mother tongue and literature	.62*** (0.01)	.49*** (0.01)	.56*** (0.01)
Grade in mathematics	.58*** (0.01)	.58*** (0.01)	.59*** (0.01)
Mean of science-related grades (physics, chemistry, biology, and geography)	.64*** (0.01)	.59*** (0.01)	.63*** (0.01)
GPA of theoretical subjects	.69*** (0.01)	.61*** (0.01)	.65*** (0.01)

Note: Standard errors are in parentheses. *** $p < .001$.

PISA were taken into account. The analyses were performed with Stata software (version 17.0).

6. Results

The results showed a clear relationship between PISA performance and school achievement in both the PISA 2015 and PISA 2018 data (Table 2). Regarding the correlations between PISA proficiency scores and corresponding school grades, all these correlations were around 0.60 and statistically significant ($p < .001$). The highest correlation ($r = 0.66$) was between proficiency scores in science and science-related grades in the PISA 2015 data. PISA proficiency scores also correlated with non-corresponding school grades, and in some cases even somewhat stronger than with corresponding school grades. The correlations between PISA proficiency scores and GPA of theoretical subjects ranged from 0.61 to 0.69. GPA correlated weakest with PISA mathematics scores. These results indicated that PISA proficiency in different domains is not discriminatory but reflects a wide range of school achievement.

In addition to correlations between PISA proficiency scores and school achievement, we also investigated the extent to which PISA proficiency levels predict corresponding school grades and whether the correspondence changes when students' gender and SES are controlled for. Table 3 shows the results of the linear regression analysis. In Model 1, only the PISA proficiency levels were included. The results showed that PISA proficiency levels predict grades. The results were quite similar for both the PISA 2015 and PISA 2018 data. For students at Level 1, the mean of the corresponding grade was around 6.7 in reading, 6.5 in mathematics, and 6.6 in science.

Table 3
Predictors of grades.

	PISA 2015		PISA 2018	
	Model 1	Model 2	Model 1	Model 2
	B (SE)	B (SE)	B (SE)	B (SE)
Mother tongue and literature				
PISA proficiency level in reading				
Level 1	Reference	Reference	Reference	Reference
Level 2	0.39*** (0.07)	0.27*** (0.08)	0.51*** (0.07)	0.36*** (0.06)
Level 3	1.05*** (0.08)	0.80*** (0.08)	1.12*** (0.06)	0.83*** (0.06)
Level 4	1.67*** (0.08)	1.30*** (0.08)	1.74*** (0.06)	1.35*** (0.06)
Level 5	2.15*** (0.08)	1.68*** (0.08)	2.23*** (0.06)	1.74*** (0.06)
Level 6	2.56*** (0.12)	2.02*** (0.13)	2.68*** (0.10)	2.08*** (0.10)
Gender				
Girl		Reference		Reference
Boy		−0.76*** (0.03)		−0.75*** (0.03)
ESCS		0.25*** (0.02)		0.24*** (0.02)
Constant	6.72*** (0.07)	7.31*** (0.07)	6.73*** (0.05)	7.32*** (0.05)
R^2	.33	.45	.37	.48
N	5013	4959	4744	4682
Mathematics				
PISA proficiency level in mathematics				
Level 1	Reference	Reference	Reference	Reference
Level 2	0.46*** (0.09)	0.38*** (0.09)	0.43*** (0.08)	0.34*** (0.08)
Level 3	1.16*** (0.09)	1.03*** (0.09)	1.18*** (0.07)	1.01*** (0.07)
Level 4	1.89*** (0.08)	1.72*** (0.08)	1.94*** (0.07)	1.72*** (0.07)
Level 5	2.53*** (0.09)	2.35*** (0.09)	2.52*** (0.09)	2.29*** (0.09)
Level 6	2.95*** (0.12)	2.77*** (0.12)	2.93*** (0.15)	2.67*** (0.15)
Gender				
Girl		Reference		Reference
Boy		−0.45*** (0.04)		−0.53*** (0.03)
ESCS		0.20*** (0.03)		0.26*** (0.03)
Constant	6.50*** (0.07)	6.78*** (0.07)	6.54*** (0.06)	6.87*** (0.06)
R^2	.34	.37	.33	.38
N	5015	4961	4745	4683
Science				
PISA proficiency level in science				
Level 1	Reference	Reference	Reference	Reference
Level 2	0.46*** (0.07)	0.37*** (0.06)	0.54*** (0.06)	0.44*** (0.06)
Level 3	1.15*** (0.06)	0.98*** (0.06)	1.21*** (0.05)	1.02*** (0.05)
Level 4	1.84*** (0.07)	1.62*** (0.06)	1.85*** (0.06)	1.59*** (0.06)
Level 5	2.39*** (0.07)	2.12*** (0.07)	2.38*** (0.06)	2.06*** (0.06)
Level 6	2.84*** (0.10)	2.58*** (0.10)	2.72*** (0.10)	2.37*** (0.10)
Gender				
Girl		Reference		Reference
Boy		−0.47*** (0.03)		−0.53*** (0.03)
ESCS		0.26*** (0.02)		0.29*** (0.02)
Constant	6.53*** (0.06)	6.86*** (0.06)	6.62*** (0.05)	6.98*** (0.05)
R^2	.42	.48	.38	.45
N	5014	4960	4745	4684

Note: *** $p < .001$.

When moving from one level to the next, the average grade increased some 0.50 units. The association ($R^2 = 0.42$) was strongest for science-related subjects.

In Model 2, students' gender and SES were added. After adding these variables, the coefficients of the PISA proficiency levels decreased, and R^2 values increased slightly. After controlling for gender and SES, PISA proficiency levels, however, remained related to grades. Both gender and SES also seemed to be related to grades. The coefficient of gender was negative, which means that boys had lower school grades in all three subjects. In mother tongue and literature, the effect of gender (PISA 2015: B = −.76, PISA 2018: B = −.75) was greater than in other subjects. The coefficient of SES was positive and ranged between 0.20 and 0.29. Thus, students with higher SES also had higher grades.

7. Discussion

The results of the PISA study have been widely referred to by actors in different nations' policy and governance (Baird et al., 2016), and the PISA data have been utilized as secondary data in a large number of studies (Hopfenbeck et al., 2018). Nevertheless, quite little is known about the correspondence between PISA performance and achievement in education. The purpose of this study was to examine whether PISA performance predicts school achievement as measured by grades.

The study showed that PISA proficiency scores correlated with corresponding grades and non-corresponding grades alike. The results were consistent with that of Fischbach et al. (2013) who also found that discriminant predictive power of PISA performance was weak. In this study, the correlations were higher than they were in the study by Fischbach et al. (2013) that was conducted with Luxembourg's PISA data. However, the magnitude of correlations was similar to those found in the Finnish national assessment of learning outcomes, in which students' self-reported grades were used (e.g., Lappalainen, 2004). In this study, students' grades were from the national register and were real grades from their school-leaving certificates. The PISA test is not a curriculum-based test but instead measures the knowledge and skills needed in modern society. Moreover, some claim the test only assesses the knowledge and skills needed in test situations (Dohn, 2007). Grades, again, are based on the aims of the curriculum, and they describe student achievement not only in a single situation but over the long run. Nevertheless, correspondence between the PISA and final assessment of basic education appears relatively high in Finland. A possible explanation might be that in Finland, the national core curriculum for basic education emphasizes the development of students' thinking and inquiry skills that are essential in the PISA assessment.

Previous studies have demonstrated that girls outperform boys when performance has been assessed by grades and that boys receive worse grades than could be expected by their performance on the standardized test (Klapp Lekholm & Cliffordson, 2008, 2009; Matějů & Smith, 2015; Voyer & Voyer, 2014). This study confirms these results. Boys' grades were somewhat lower than girls' grades in all three subjects. This result was the clearest in mother tongue and literacy, in which the outperformance by girls on the PISA test was also the largest (OECD, 2019c). Earlier studies (Bowers, 2011; Klapp Lekholm & Cliffordson, 2009) showed that grades are not only affected by achievement but also by non-cognitive factors such as motivation. Thus, future studies that take these non-cognitive factors into account when studying gender differences in the correspondence between grades and PISA performance will be needed.

This study indicated that PISA proficiency level is related to grades after controlling for SES. This result agrees with those obtained by Fischbach et al. (2013). However, students' SES also seemed to be related to grades. Students who had higher SES seemed to receive slightly higher grades. Some earlier studies and national assessments (e.g., Julin & Rautopuro, 2016; Klapp Lekholm & Cliffordson, 2008; Mattila, 2002) have shown that the grades of students with more highly educated parents were somewhat lower with relation to the performance shown in the assessments, which differs from the results of this study. In this study, SES describes, in addition to parents' educational level, parents' occupational status and household possession. Therefore, the results of this study and the above-mentioned earlier studies and assessments are not fully comparable.

The data used in this study involve some limitations. Rutkowski et al. (2010) highlighted a problem that relates to data that link PISA data with outside data, as we have done in this study by combining PISA data and national register of education records. According to the authors, estimates that use information from the added data might be biased since the PISA PV estimation has not taken this information into account. Therefore, the results of this study need to be interpreted with caution. Additionally, we used the ESCS index of the PISA data as a gage of students' SES. This index is based on the background information gathered from the students. It has been argued that students' self-reported background information and the socio-economic background measures in the PISA data are insufficient (Rutkowski & Rutkowski, 2013). On the other hand, it has also been shown that, of the SES variables used in the PISA study, the ESCS index is the best predictor for student achievement (Lee et al., 2019).

Notwithstanding the limitations of this study, the study suggests that PISA performance predicts achievement at school. However, further research should focus more on investigating how PISA performance predicts learning outcomes in upper secondary and tertiary education. Since the aim of the PISA study is to assess the knowledge and skills students need in further studies and for participation in society (OECD, 2016), longitudinal research is required to assess whether the PISA test is valid in this respect.

Declarations of Competing Interest

None.

References

Act on the National Registers of Education Records, Qualifications and degrees, 884. (2017). (Fin.).

Baird, J. A., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G., et al. (2016). On the supranational spell of PISA in policy. *Educational Research, 58*(2), 121–138. https://doi.org/10.1080/00131881.2016.1165410

Basic Education Act, 628. (1998). (Fin.).

Basic Education Decree, 852. (1998). (Fin.).

Bertschy, K., Cattaneo, M. A., & Wolter, S. C. (2009). PISA and the transition into the labour market. *Labour (Committee on Canadian Labour History), 23*(1), 111–137. https://doi.org/10.1111/j.1467-9914.2008.00432.x

Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation, 17*(3), 141–159. https://doi.org/10.1080/13803611.2011.597112

Dohn, N. B. (2007). Knowledge and skills for PISA — Assessing the assessment. *Journal of Philosophy of Education, 41*(1), 1–16. https://doi.org/10.1111/j.1467-9752.2007.00542.x

Eurydice. (2021). National education systems. https://eacea.ec.europa.eu/national-policies/eurydice/national-description_en.

Finnish National Agency for Education. (2004). Perusopetuksen opetussuunnitelman perusteet 2004. [The National Core Curriculum for Basic Education 2004]. https://www.oph.fi/sites/default/files/documents/perusopetuksen-opetussuunnitelman-perusteet_2004.pdf.

Finnish National Agency for Education. (2020). Perusopetuksen päättöarvioinnin kriteerit [The criteria for final assessment of basic education]. The Rule of Finnish National Agency for Education 5042/2020. https://www.oph.fi/sites/default/files/documents/Perusopetuksen%20p%C3%A4%C3%A4tt%C3%B6arvioinnin%20kriteerit%2031.12.2020_0.pdf.

Fischbach, A., Keller, U., Preckel, F., & Brunner, M. (2013). PISA proficiency scores predict educational outcomes. *Learning and Individual Differences, 24*, 63–72. https://doi.org/10.1016/j.lindif.2012.10.012

Harju-Luukkainen, H., Vettenranta, J., Ouakrim-Soivio, N., & Bernelius, V. (2016). Differences between students' PISA reading literacy scores and grading for mother tongue and literature at school: A geostatistical analysis of the Finnish PISA 2009 data. *Education Inquiry, 7*(4), 463–479. https://doi.org/10.3402/edui.v7.29413

Hildén, R., & Rautopuro, J. (2017). On an equal footing? Comparing school grade and national evaluation results in Finland. In N. Pyyry, L. Tainio, K. Juuti, R. Vasquez, & M. Paananen (Eds.), *Changing subjects, changing pedagogies: Diversities in school and education* (pp. 242–259). Publications of the Finnish Research Association for Subject Didactics. Studies in Subject Didactics 13.

Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J. A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research, 62*(3), 333–353. https://doi.org/10.1080/00313831.2016.1258726

Hopfenbeck, T. N., & Maul, A. (2011). Examining evidence for the validity of PISA learning strategy scales based on student response processes. *International Journal of Testing, 11*(2), 95–121. https://doi.org/10.1080/15305058.2010.529977

Hotulainen, R., Rimpelä, A., Hautamäki, J., Karvonen, S., Kinnunen, J.M., Kupiainen, S. et al. (2016). *Osaaminen ja hyvinvointi yläkoulusta toiselle asteelle: Tutkimus metropolialueen nuorista* [Learning and well-being: Transition from lower to upper secondary education in the Helsinki Metropolitan Area]. Tutkimuksia [Studies] 398. University of Helsinki.

Jakku-Sihvonen, R. (2013). Oppimistulosten arviointijärjestelmistä ja niiden kehittämishaasteista [On assessment systems for learning outcomes and developmental challenges of these systems]. In A. Räisänen (Ed.), *Oppimisen arvioinnin kontekstit ja käytännöt [Contexts and practices for the assessment of learning]* (pp. 13–36). Finnish National Agency for Education.

Jakku-Sihvonen, R., & Kuusela, J. (2002). *Mahdollisuuksien koulutuspolitiikan tasa-arvo [Equal opportunities in education policy]. Arviointi [Evaluation] 7/2002.* Finnish National Agency for Education.

Julin, S., & Rautopuro, J. (2016). *Läksyt tekijäänsä neuvovat. perusopetuksen matematiikan oppimistulosten arviointi 9. vuosiluokalla 2015* [You learn as you do your homework. the assessment of learning outcomes in mathematics on the 9th grade in 2015]. julkaisut [Publications] 20/2016. Finnish Education Evaluation Centre.

Kivinen, O., & Hedman, J. (2017). Moniselitteiset PISA-tulokset ja niiden ongelmalliset koulutuspoliittiset tulkinnat [Ambiguous PISA results and their problematic interpretations in education policy]. *Politiikka, 59*(4), 250–263.

Klapp Lekholm, A., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: Effects of gender and family background. *Educational Research and Evaluation, 14*(2), 181–199. https://doi.org/10.1080/13803610801956663

Klapp Lekholm, A., & Cliffordson, C. (2009). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation, 15*(1), 1–23. https://doi.org/10.1080/13803610802470425

Lappalainen, H.P. (2004). *Kerroin kaiken tietämäni – perusopetuksen äidinkielen ja kirjallisuuden oppimistulosten arviointi 9. vuosiluokalla 2003* [I told everything i knew. the assessment of learning outcomes in mother tongue and literature on the 9th grade in 2003]. Oppimistulosten arviointi [Assessment of learning outcomes] 2/2004. Finnish National Agency for Education.

Lappalainen, H.P. (2006). *Ei taito taakkana ole – perusopetuksen äidinkielen ja kirjallisuuden oppimistulosten arviointi 9. vuosiluokalla 2005* [A skill in not a burden. the assessment of learning outcomes in mother tongue and literature on the 9th grade in 2005]. Oppimistulosten arviointi [Assessment of learning outcomes] 1/2006. Finnish National Agency for Education.

Lee, J., Zhang, Y., & Stankov, L. (2019). Predictive validity of SES measures for student Achievement. *Educational Assessment, 24*(4), 305–326. https://doi.org/10.1080/10627197.2019.1645590

Matějů, P., & Smith, M. L. (2015). Are boys that bad? Gender gaps in measured skills, grades and aspirations in Czech elementary schools. *British Journal of Sociology of Education, 36*(6), 871–895. https://doi.org/10.1080/01425692.2013.874278

Mattila, L. (2002). *Perusopetuksen matematiikan oppimistulosten kansallinen arviointi 9. vuosiluokalla 2002* [National assessment of mathematics learning outcomes in the ninth grade in basic education 2002]. oppimistulosten arviointi [Assessment of learning outcomes] 8/2002. Finnish National Agency for Education.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA). https://timssandpirls.bc.edu/isc/publications.html.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA). https://timssandpirls.bc.edu/isc/publications.html.

Murdoch, J., Kamanzi, P. C., & Doray, P. (2011). The influence of PISA scores, schooling and social factors on pathways to and within higher education in Canada. *Irish Educational Studies, 30*(2), 215–235. https://doi.org/10.1080/03323315.2011.569142

OECD. (2009). *PISA data analysis manual: SPSS* (2nd ed.). OECD Publishing. https://doi.org/10.1787/9789264056275-en

OECD. (2016). *PISA 2015 results: Vol. 1 excellence and equity in education*. OECD Publishing. https://doi.org/10.1787/9789264266490-en

OECD. (2019a). *PISA 2018 assessment and analytical framework*. a. OECD Publishing. https://doi.org/10.1787/b25efab8-en.

OECD. (2019b). *PISA 2018 results: Vol. 1 what students know and can do*. b. OECD Publishing. https://doi.org/10.1787/5f07c754-en.

OECD. (2019c). *PISA 2018 results: Vol. 2 where all students can succeed*. c. OECD Publishing. https://doi.org/10.1787/b5fd1b8f-en.

Ouakrim-Soivio, N., Rautopuro, J., & Hildén, R. (2018). Shadows under the North Star: The inequality developing in Finnish school education. *Nordidactica: Journal of Humanities and Social Science Education, 3*, 65–86.

Rutkowski, D., & Rutkowski, L. (2013). Measuring socioeconomic background in PISA: One size might not fit all. *Research in Comparative and International Education, 8*(3), 259–278, 10.2304%2Frcie.2013.8.3.259.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142–151, 10.3102%2F0013189X10363170.

Rutkowski, L., & Rutkowski, D. (2010). Getting it 'better': The importance of improving background questionnaires in international large-scale assessment. *Journal of Curriculum Studies, 42*(3), 411–430. https://doi.org/10.1080/00220272.2010.487546

Rutkowski, L., & Rutkowski, D. (2016). A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher, 45*(4), 252–257, 10.3102%2F0013189X16649961.

Rutkowski, L., & Rutkowski, D. (2018). Improving the comparability and local usefulness of international assessments: A look back and a way forward. *Scandinavian Journal of Educational Research, 62*(3), 354–367. https://doi.org/10.1080/00313831.2016.1261044

Schleicher, A. (2007). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess? *Journal of Educational Change, 8*(4), 349–357. https://doi.org/10.1007/s10833-007-9042-x

Skolverket. (2022). Nationella prov i grundskolan [National exam in basic education]. https://www.skolverket.se/undervisning/grundskolan/nationella-prov-i-grundskolan.

Thorsen, C., & Cliffordson, C. (2012). Teachers' grade assignment and the predictive validity of criterion-referenced grades. *Educational Research and Evaluation, 18*(2), 153–172. https://doi.org/10.1080/13803611.2012.659929

Volmari, K. (2019). (Ed.) Basic education in the Nordic region. Similar values, different policies. Reports and surveys 2019:4. Finnish national agency for education. https://www.oph.fi/sites/default/files/documents/30955454_basic_education_in_the_nordic_region_netti_19_04_08_0.pdf.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204. https://doi.org/10.3758/BF03194024